ISSN 2229-5518

Speech enhancement by a Kalman filter based smoother in white and colored noise

Srilakshmi Garaga , O.Sudhakar

Abstract ___ **Since speech is non-stationary, it is common practice to segment the speech into short overlapped frames and assume the signal to be stationary with in each frame. This is known as quasi stationary assumption. The main idea in this project is to recover the clean speech signal from a sample corrupted with background noise through a telephone conversation. To achieve goal to estimate the speech spectrum using LP method and using SWLP, an improved algorithm .with this method, choosing correctly the parameters can obtain an improved robustness against the background noise. After this, the idea is to use the Kalman filtering as a tool to estimate the future clean samples from the first one in an iterative way.**

This paper presents a new approach to speech enhancement based on Kalman algorithms This approach differs from the single-channel speech enhancement. However, when carrying out Kalman smoothing, the computational cost and the data storage requirements are two specific problems. In this paper a filter-based smoother is proposed and used in the framework of speech enhancement. The algorithm is evaluated by considering a speech signal embedded in a white Gaussian noise. Simulation Results show that the proposed algorithm provides a higher improvement of signal to noise ratio (SNR) than the Kalman filtering.

Keywords __ **Speech enhancement, Kalman filter, LP, SWLP, white noise, colored noise.**

1 INTRODUCTION

The objective of speech enhancement is to improve the quality of a speech signal, often degraded by some type of distortion (for example communication channel distortion, additive noise, convolution filtering operation, etc.).The difficulty from the background noise is not useful to keep the communication if the noise of the subway, train, car, etc., prevents us from understanding.

That is the goal of this paper: to eliminate the maximum background noise keeping the quality of the voice to transmit, because there exists a limit where the voice stops to sound like human voice, it loses the nuances and starts to seem like something artificial. For this purpose we use the Kalman filter. For the experimental part , use the database form sentences recorded in English, which have been filtered by IRS filters to simulate the telephone handsets and later, it has been added the background noise. Then develop this in two environments, with white noise and with colored noise; the latter is closer to the real life noise. For each scenario, two methods will be used to model the signals: Linear Predictive Coding (LPC) and Stabilized weighted linear prediction (SWLP). Both predict the coefficients of the signals.

- *Author G.Srilakshmi is currently pursuing M.Tech in DECS, Dept of E.C.E Godavari Institute of Engineering & Technology, Rajahmundry, India. srilakshmi1853@gmail.com*
- *Co- author O.Sudhakar is working as the Assistant Professor in the Dept of E.C.EGodavari Institute of Engineering & Technology, Rajahmundry, India.*

The clean speech, denoted s(k) is often represented by a linear prediction (LP) model [1] ; in addition, the background noise is usually assumed stationary and its second order statistics can be estimated during the periods of silence, between utterances. In [2],P aliwal etal. propose a Kalman filter based speech enhancement, in which the LP parameters are estimated directly from the clean speech, supposed to be available. However, this approach cannot be computed in practice. For this reason, an Expectation-Maximization (EM) [3][4] can he considered; this is an iterative likelihood maximization method used when it is difficult to obtain a direct maximum likelihood (ML) estimate, operating in two steps (E and M steps). In [5][6], the Estep consists in carrying out a Kalman filter. It should be noted that a Kalman smoother [7][8] can also be used since it improves the estimation precision and hence weakens the residual noise in the enhanced speech. However, when completing standard Kalman smoothing [9], data storage space must be taken into account. This comprises the Kalman gain matrix, filtered mean vector and covariance matrix, one-step-forward predicted mean vector and covariance matrix. These quantities are calculated by the Kalman filter and used for backward in time recursive calculation. So, due to the enormous storage requirements and the computational cost, applying Kalman smoothing is confined to very specific applications.

Filtering algorithms are time-domain methods that attempt to either remove the noise component (Wiener filtering) or estimate the noise and speech components by a filtering approach (Kalman filtering). There is an important algorithm for speech enhancement which belongs to the group of parametric methods where the speech signal is modeled as an autoregressive process embedded in Gaussian noise.

Speech enhancement algorithms belonging to this category consist of two steps:

- Estimation of the AR coefficients and noise variances.
- Application of the Kalman filtering using the estimated parameters to estimate the clean speech from a sample of the noisy signal.

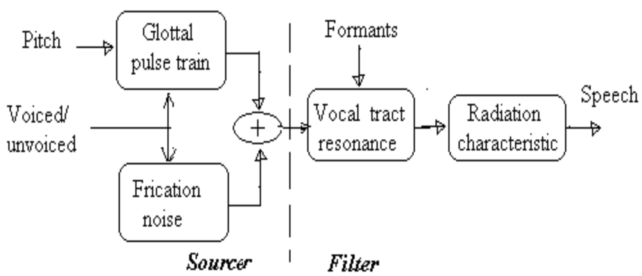


Fig. Production model voice

2 KALMAN FILTER FOR SPEECH ENHANCEMENT

2.1 Accurate estimation of LPC and SWLP

Linear prediction is one of the most powerful tools used, where a signal yn is the output of a system considering the unknown signal xn as the input with the relation,

$$y_n = -\sum_{k=1}^{p} a_k \cdot y_{n-k} + G \cdot \sum_{l=0}^{q} b_l \cdot x_{n-l} \;, \quad b_0=1$$

Where G are the parameters of a hypothesized system

The estimation of model parameters can be derived in the time domain and in the frequency domain. In general, input signal predict it, x_n as a linear weight combination of the past samples,

$$\hat{y}_n = -\sum_{k=1}^{p} a_k \cdot y_{n-k}$$

Where a_k are the predictor coefficients, p is the model order and the minus sign is for convenience. The error with this method.

$$e_n = y_n - \hat{y}_n = y_n + \sum_{k=1}^{p} a_k \cdot y_{n-k} = \sum_{k=0}^{p} a_k \cdot y_{n-k}$$

Where yn is the original signal, a0=1 and en is called residual.This measures the quality of the predictor. If denote the total squared error by E, where

$$E = \sum_n e_n^2 = \sum_n \left(y_n + \sum_{k=1}^{p} a_k \cdot y_{n-k} \right)^2$$

Then, to minimize E:

$$\frac{\partial E}{\partial a_i} = 0, \quad 1 \le i \le p$$

This method is called method of least squares and the parameters ak are calculated as a result of the minimization of the mean or total squared error with respect to each of the parameters (it is called autocorrelation criterion too).
With the auto-correlation method,

$$\sum_{k=1}^{p} a_k \cdot R(i-k) = -R(i), \quad 1 \le i \le p$$

$$E_p = R(0) + \sum_{k=1}^{p} a_k \cdot R(k)$$

Where,

$$R(i) = \sum_{n=-\infty}^{\infty} y_n \cdot y_{n+1}$$

Is the auto-correlation function of yn. Then, observe that: R (− i) = R (i) .Like the coefficients R (i-k) form an auto-correlation matrix
The coefficient vector a = (a0 , a1, … ap) T, of a FIR predictor with order p, which minimizes the prediction error energy. The corresponding all-pole filter is obtained as H(z) = 1 / A(z), where A(z) is the z-transform of a. To achieve this, it exists a formula to modify the weight function of WLP and, in this way, to reach the stability of the all-poles filter.

All of this can be carried out by changing the elements of the secondary diagonal of the B matrix:

$$B_{i,i+1} = \begin{cases} \sqrt{w_i / w_{i-1}} \;, & \text{if } w_i \le w_{i+1} \\ 1, & \text{if } w_i > w_{i+1} \end{cases}$$

From now on, the WLP method calculated using the B matrix, is called stabilized weighted linear prediction (SWLP), where the stability of the all poles filter is guaranteed. The main concept in WLP, is the time domain weight function. Choosing an appropriate waveform, one can temporally emphasize or attenuate the weight of the residual energy prior to the optimization of the filter parameters. The weight function was chosen basing on the short-time energy (STE),

$$w_n = \sum_{i=0}^{M-1} x_{n-i-1}^2$$

Where M is the length of the window.

2.2 The process to be estimated:

The Kalman filter has the goal of solving the general problem of estimate Rm of a process controlled in discrete time, which is the state X dominated by a linear equation in stochastic difference in the following way:

$$X_n = A \cdot X_{n-1} + w_{n-1}$$

With a measure $Y \in R^n$, that is:

$$Y_n = C \cdot X_n + v_n$$

The random variables wn and vn represent the process and

ISSN 2229-5518

the measure error, respectively. It is assumed they are independent of each other, and are white noise variables with normal probability distribution:

$$p(w) \approx N(0, R_w)$$
$$p(v) \approx N(0, R_v)$$

In practice, the covariance matrix of the process's perturbation, Rw, and the measure's perturbation, Rv, could change in time, but for simplicity, it is assumed they are constants.

2.3 The Kalman filter and the state- space notation

The Kalman filter is the main algorithm to estimate dynamic systems specified with the state-space model. Actually, the state-space models and the Kalman filter models are often used as synonymous. The estimation and control of the problems of this methodology are based on stochastic models, assuming errors in the measures. The performance of the state-space model for a linear system captures a yn vector with nx1 order associated to an unknown xn vector with mx1 order, known as state vector.
In speech processing, I assume the case with a signal received by a single microphone and additive noise. Let the signal measured by the microphone be given by:

$$y_n = x_n + v_n$$

Where yn is the observed signal, xn is the desired input and vn is the additive background noise (zero-mean noise). Furthermore, like xn is modeled as autoregressive, assume the standard LPC modeling for the speech signal over an analysis frame:

$$x_n = -\sum_{k=1}^{m} a_k \cdot x_{n-k} + w_n = -a_n * x_{n-1} + w_n$$

On the other hand, the last equation can be reformulated in a state space presentation with the state transition matrix or companion matrix:

$$A_{n+1,n} = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{m-1} & -a_m \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

Then, write the state-space form:

$$x_{n+1} = A_{n+1,n} \cdot x_n + u \cdot w_{n+1}$$
$$y_n = C \cdot x_n + v_n$$

Where C is a matrix of the system, Wn+1 is the noise indoor and vn is the noise outdoor.

The first of these equations is known as process equation and the second one as measurement equation. The first equation shows the relation among previous states and futures states, while the second one gives us the correspondence between the internal state of the system and how it can be observed [10]. These equations are useful for most of the linear estimation methods, like the Kalman filter described above.

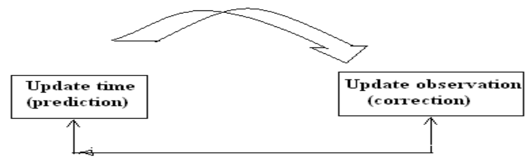


Figure 2: The kalman filter cycle

2.4 First option: assuming white noise:

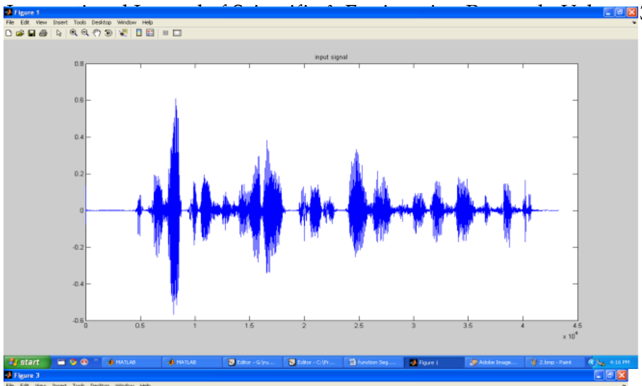
First consider white noise; the coefficients of the noise added signal are calculated through LPC method and use this estimation for all the process, assuming they don't change.

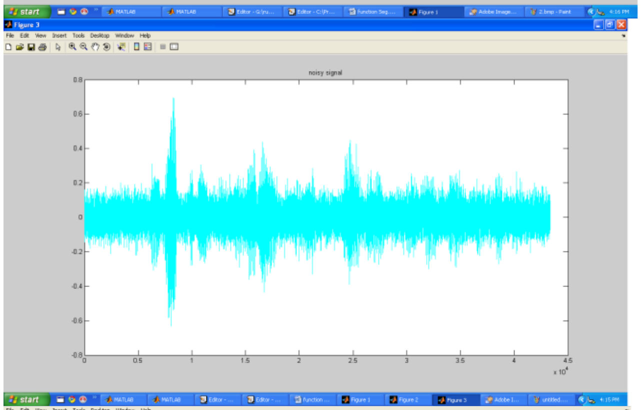
2.5 Second option: assuming colored noise

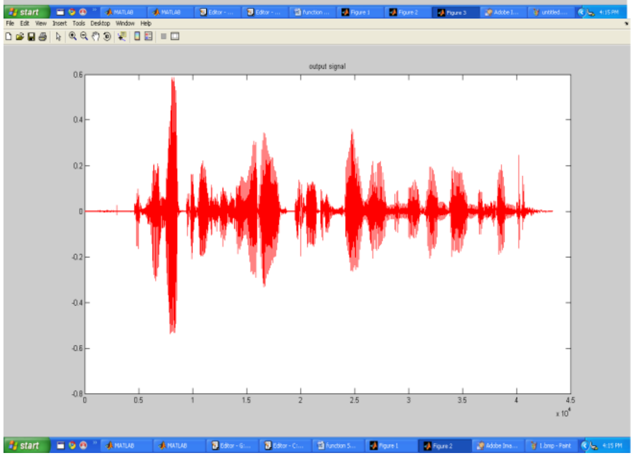
The only difference in the code with this change is that the coefficients of the noise added signal change with the time. For that reason, inside the loop, each time calculate the coefficients of the original signal with noise, then calculate the coefficients of the noise signal too; here LPC method is used for all the tries.

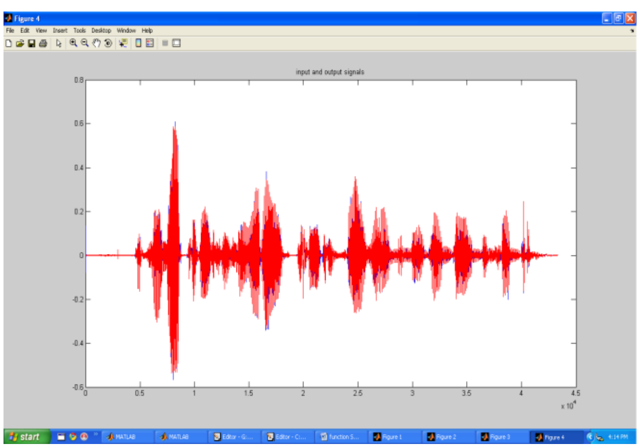
3. RESULTS

Fig 3.a Input Signal
Fig 3.b Noisy Signal
Fig. 3.c Output Signal
Fig. 3.d Input and Output Signal









4. CONCLUSION

The Kaman Filter has been widely used in many areas from tracking to speech enhancement. Since the speech signal is often assumed stationary during an analysed frame (20-30 ms), the Kalman smoother can he carried out and provides better estimates of the state since it is based on a higher number of observations. However, the enormous storage requirements of the Kalman smoother coniine its applications in practice. In this paper, our purpose was to present an alternative to the Kalman smoother,filter-based smoother (FBS), which does not require such a storage space for the vectors and matrices during the Kalman filtering process. The evaluation of the FBS-based speech enhancement algorithm. The results show that the proposed FBS based algorithm can provide a higher SNR improvement than the KF based algorithm.

REFFERENCES

[1]. J. R. Deller, J. H. L. Hanson and J. G.Proakis. Diccrere-The Processing of Speech Signal, EEE PRESS, New York 2000.

[2]. K. K. Paliwal and A.Basu, “A speech enhancement method based on Kahan filtering, Proceedings of ICASSP’87,pp.177-180. Dallas, TX, USA, 1987.

[3]. T.K.Mmn, “The Expectation-Maxlmizaion Algorithm”,IEEE Signof Processing Magazine, pp. 47-60, Nov. 1996.

[4]. S. V. Vaseghi, Advanced Signal Processing and Digital Noise Reducrion, Wiley Teubner, 1996.

[5]. M. Gabrea, E. Gnvel and M. Najim, “A Single Microphone Kalman Filter-Based Noise Canceller”, IEEE Signal Pmcessing Letters, march 1999, pp. 53-55

[6]. J. D. Gibson, B. Koa and S. D. Gny, “Filtering of Colored Noise for Speech Enhancement and Coding”, IEEE Trans. onSignalPmces~ingv,o l. 39,pp. 1732.1742, August 1991.

[7]. J.C asals, M. Jerez and S . Sotoca, “Exact Smoothing for Stationary and Nonstationary Time Series”, International Joumal ofForecasting, vo1.16, pp.59-69,2000,

[8]. S. Cannot, D. Burchtein and E. Weinstein, “Iterative and Sequential Kalamn Filter-Based Speech Enhancement algorithms”, IEEE Trans. on Speech and audio Proce.rsing. pp.373-385, July 1998

[9]. J. Vermaak, M. Niranjan and S. J. Godsill, “Compaing Solution Methodologies for Speech Enhancement within a Bayesian Framework, Technical Repon CUED/FINFENGRR329, Cambridge University Engineering Department, August 1998.

[10]. Sadaoki Furui, “Digital speech processing, synthesis, and recognition”, Marcel Dekker, 2001